

CFN: A Complex-valued Fuzzy Network for Sarcasm Detection in Conversations

Yazhou Zhang¹, Yaochen Liu², Qiuchi Li³, Prayag Tiwari⁴, Benyou Wang⁵, Yuhua Li, Hari Mohan Pandey⁶, Peng Zhang and Dawei Song⁷

Abstract—Sarcasm detection in conversation (SDC), a theoretically and practically challenging artificial intelligence (AI) task, aims to discover elusively ironic, contemptuous and metaphoric information implied in daily conversations. Most of the recent approaches in sarcasm detection have neglected the intrinsic vagueness and uncertainty of human language in emotional expression and understanding. To address this gap, we propose a complex-valued fuzzy network (CFN) by leveraging the mathematical formalisms of quantum theory (QT) and fuzzy logic. In particular, the target utterance to be recognized is considered as a quantum superposition of a set of separate words. The contextual interaction between adjacent utterances is described as the interaction between a quantum system and its surrounding environment, constructing the quantum composite system, where the weight of interaction is determined by a fuzzy membership function. In order to model both the vagueness and uncertainty, the aforementioned superposition and composite systems are mathematically encapsulated in a density matrix. Finally, a quantum fuzzy measurement is performed on the density matrix of each utterance to yield the probabilistic outcomes of sarcasm recognition. Extensive experiments are conducted on the MUSTARD and the 2020 sarcasm detection Reddit track datasets, and the results show that our model outperforms a wide range of strong baselines.

Index Terms—Sarcasm detection, emotion recognition, fuzzy logic, quantum theory, artificial intelligence.

I. Introduction

SARCAISM can be traced back to ancient Greece, and was first recorded in English in 1579 [1]. It is a kind

Manuscript received October 19, 2020; revised February 15, 2021. This work is supported by National Science Foundation of China under grant No. 62006212, the Doctoral Scientific Research Foundation of Henan Province (grant No. 0215-13501050028, 2020BSJJ031), the Natural Science Foundation of China (grant No. 2018YFC0831704, 62072414, U1636203, 61772363, U1736103, U1504608, 61802352), the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 721321. (Corresponding authors: Dawei Song and Hari Mohan Pandey)

Y. Zhang and Y. Li are with Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou 450002, China (email: yzzhang@zzuli.edu.cn).

Q. Li, P. Tiwari, and B. Wang are with the Department of Information Engineering, University of Padua, Padua, Italy.

P. Tiwari is with the Department of Computer Science, Aalto University, Finland (email: prayag.tiwari@iee.org).

H. M. Pandey is with the Department of Computer Science, Edge Hill University, Ormskirk, United Kingdom (email: pandeyh@edgehill.ac.uk).

P. Zhang is with the Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin 300350, China.

Y. Liu and D. Song are with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China, and D. Song is also with the School of Computing and Communications, The Open University, Walton Hall, United Kingdom (email: dawei.song2010@gmail.com).

Y. Zhang and Y. Liu contribute equally.

of rhetorical strategy that is intended to express criticism or mock emotions by means of hyperbole, figuration, etc [2], [3]. Through the theory discussion for a long time, it is finally defined as “a mode of satirical wit depending for its effect on bitter, caustic, and often ironic language that is usually directed against an individual” [4]. The recent advancement of Internet and social network services has led to a huge and increasing usage of ironic language, which plays an important role in daily discourse. We here give two real-life examples: (1) a waiter sees one client struggling to open a door and asks the client, “Do you want help?”, if the client replies by saying, “No, I’m really enjoying the challenge”. Then the waiter knows he’s being sarcastic. (2) When a husband comes home after a long day at work, he expresses his sarcastic attitude, “I love working 40 hours a week, well done!”.

Thus, identifying the sarcasm emotion of user-generated texts has a large potential for a wide range of domains, e.g., to help manufacturers predict the attitudes of consumers toward their products and to help political associations understand general public opinions. For example, Donald Trump takes an irony tone at Joe Biden in his tweets such as “He is actually somewhat better than a rabid dog”, due to his stance against Biden who appeared to be his new rival in the next presidency election. As another example, the third-party sellers on Amazon want to find public or consumer opinions and emotions about their products and services. Hence, there has been an increasing interest from both academia and industry in detecting sarcasm in texts [5], [6].

An effective sarcasm detector is also beneficial to applications like sentiment analysis [7], humor analysis [8], brand management [9], business intelligence and more broadly across our daily lives [10]. Sarcasm detection refers to the use of natural language processing (NLP), statistics and machine/deep learning methods to recognize sarcasm or irony orientations for various granularities of texts at the sentence, document or conversation levels. It is often formalized as the binary classification problem [11]. Previous sarcasm detection methods in the literature have mainly focused on analyzing narrative texts, e.g., product reviews, tweets, etc., without involving interaction among the writers or speakers.

Currently, there are a series of emerging conversational sarcasm detection models that target at detecting the sarcastic attitude of multiple speakers in an conversation. Compared with the traditional sarcasm detection, SDC is more challenging for two reasons: (1) in the conversation, the attitude of each speaker is heavily influenced by other speakers, thus they are inseparable and cannot be treated independently; (2) the



Fig. 1: Sample sarcastic utterance in a conversation along with its context.

interaction among people carries a wealth of information, such as their social relationships, stances, etc. However, such models, including the state of the art [12] are focused on investigating the role of conversation context or learning the contextual dependencies. They have not yet taken into consideration the inherent vagueness and uncertainty of human language in sarcasm expression, which needs to be studied from a more general cognitive science perspective. For illustration, Figure 1 provides a sarcastic example from the MUSTARD dataset [2].

In cognitive science, emotions are considered as the uncertain and vague part of human perception [13]. The uncertainty mainly refers to the spontaneity of emotional activities, where emotions are generated automatically without any rational reasoning process, and the change in emotional states does not involve any rational logical reasons. Even we have collected all prior knowledge, we might not determine emotional states. The vagueness refers that there is not a sharp line between different emotions, e.g., sad and depressed. Since emotion is a positive or negative personal experience, it cannot be as clear and determinate as rational logic.

As a specific form of emotional expression, sarcasm naturally inherits these characteristics. To model the fuzziness, a large body of fuzzy logic based models has been proposed [14], [15], [16], [17], [18]. A detailed literature review is given in Sec. II-B. They usually extract the syntactic and semantic features in a sentence by designing machine/deep learning architectures and obtain the predictions through using various fuzzy membership functions or constructing applicable *if-then* rules. Most of them neglect another key factor determining the sarcasm polarity, namely the uncertainty of human language.

In recent years, quantum theory (QT), as a mathematical formalism to model the uncertain particle behaviors in quantum physics, has been adopted for describing elusive human cognitive and emotional activities in various AI tasks [19], [11]. For instance, the quantum language model (QLM) [20] and neural network-based QLM [21] represented user's information needs and documents as density matrices (DMs) in a common quantum probabilistic space. The quantum sentiment repre-

sentation (QSR) model [22], [23] learned both the sentiment and semantic information with an improved version of QLM. However, such QT-based models are limited in that they restrict the models to finite vector spaces over real numbers. The potential for complex-valued formulations has not been fully developed. To address this problem, Wang and Li [24] defined a complex semantic Hilbert space to capture the “quantumness” in the cognitive aspect of human language. Nonetheless, their model randomized the complex phase instead of digging into its concrete meaning, and did not take into account the contextual interaction information, which is crucial for understanding human language.

In this paper, we argue that there are some fundamental connections between QT and fuzzy logic, since the fuzzy logic interpretation of quantum mechanics has been demonstrated under some circumstances [25]. Hence, unifying the quantum theory formalism and fuzzy logic would give us a more powerful theoretical framework to capture the subtle sentiments and semantics behind multiparty conversations. We thus propose a complex-valued fuzzy network, termed CFN, to jointly capture the uncertainty and vagueness of human language in sarcastic expression¹. To model the uncertainty, each utterance is treated as a quantum superposition of a set of basis words, which is represented by a complex-valued vector, where each component adopts an amplitude-phase form $z = re^{i\theta}$. The contextual interaction between adjacent utterances is described as the interaction between a quantum system and its surrounding environment, constructing a composite system. To model the fuzziness, the weight of interaction is determined by the fuzzy membership function. Then, the speaker's sarcastic attitude is viewed as a quantum mixed system composed of composite systems, which is mathematically encapsulated in a density matrix. Finally, considering the fact that all the information contained in one system (which, in this paper, refers to each utterance) is represented by the probability distribution of quantum measurement results, sarcastic features are extracted via the concept of quantum measurement, which is a natural choice. A fuzzy quantum measurement is performed on the density matrix of each target utterance to extract the probabilistic features, while these features are passed to a fully connected softmax layer to yield predictions over the sarcasm labels.

We have designed and carried out extensive experiments on two benchmark conversational sarcasm datasets, i.e., MUSTARD and the 2020 sarcasm detection Reddit track, to demonstrate the effectiveness of the proposed CFN framework in comparison with a wide range of baselines, including a machine learning approach (i.e., support vector machine, SVM) and seven state-of-the-art sarcasm detection approaches (i.e., convolutional neural network (CNN), bidirectional gated recurrent unit (BiGRU), multi-head attention-based bidirectional long-short memory (MHA-BiLSTM) network, bidirectional encoder representations from transformers (BERT), RCNN-RoBERTa, contextual sarcasm detection network (C-Net) and a multi-task learning (MTL) framework). The results show that the CFN

¹We have made our source code publicly available at <https://github.com/whatc0de/CFN-maste>

significantly outperforms a wide range of comparative models.

The rest of this paper is organized as follows. Section II outlines the related work. Section III introduces the preliminaries of quantum theory and fuzzy logic. In Section IV, we describe the proposed complex-valued fuzzy network framework in detail. In Section V, we report the empirical experiments and analyze the results. Section VI concludes the paper and points out future research directions.

II. Related Work

A. Sarcasm Analysis

Sarcasm is a very subtle form of metaphorical language, where the literal meaning of the sentence is contrary to its true interpretation. In NLP, sarcasm detection is typically treated as a text classification task. Generally speaking, there exist three categories of approaches in the current literature: rule-based, machine learning-based and deep learning-based approaches.

Rule-based approaches. The rule-based approaches infer the overall sarcasm polarity of a piece of text based on refined sarcasm rules, which do not require a large data corpus and training algorithms. The early research in this direction assumed interjections as stereotypic of sarcastic text. Bharti et al. [26] proposed two lexicon based approaches, one of which is a parsing-based lexicon generation algorithm (PBLGA) and the other is based on the occurrence of the interjection word. Hernandez et al. [27] used the semantic relatedness between words as the sarcastic features. Bouazizi et al. [28] defined four sets of features that cover different types of sarcasm, and used these features to classify sarcastic tweets. Clews and Kuzma [29] applied a string matching strategy against positive sentiment and used interjection lexicons to judge sarcasm. As the rule-based approaches largely depend on rules and patterns, their classification accuracy is generally lower than machine/deep learning approaches. Satoshi Hiai [30] extracted sarcastic sentences in product reviews using classification rules and classified the sentences into eight classes by focusing on evaluation expressions. Kamal et al. [31] proposed a self-deprecating sarcasm detection approach using an amalgamation of rule-based techniques.

The rule-based approaches do not consider the contextual interaction. Since they rely heavily on sarcasm rules or lexicons, their classification performance is generally inferior to machine/deep learning-based approaches.

Machine learning-based approaches. These methods mainly make use of machine learning methods, such as random forest, support vector machines, and neural networks. They often involve building classifiers from labeled data, essentially a supervised classification task. For instance, Lunando et al. [32] employed the negativity information and the number of interjection words in the translated SentiWordNet as features, and fed them into various machine learning classifiers. Habernal et al. [33] evaluated two machine learning classifiers with various combinations of features, e.g., N-gram, POS, etc., on both the Czech and English sarcasm datasets. Mukherjee and Bala [34] tested a range of feature sets using the Naive Bayes and fuzzy clustering algorithms for sarcasm detection of online text. Sharma [35] used features of user's account and tweets,

and devised three machine learning algorithms for the task of potential rumour origin detection. Kumar and Garg [36] compared the performance of several machine learning algorithms, including support vector machines, decision trees, and random forest, etc.

There have been a few sarcasm detection approaches that explored contextual features to acquire shared knowledge between the speakers. Rajadesingan et al. [37] used the user's past tweets to construct a behavioral modeling framework tuned for detecting sarcasm. Joshi et al. [38] proposed a sequence labeling approach and showed that the history utterances help improve the performance of sarcasm detection.

There are a range of machine learning-based rumor propagation and recognition approaches that also develop similar strategy to investigate rumor. Belen and Pearce [39] checked general initial conditions of ignorants, spreaders and stiflers, and analyzed how the initial conditions bear on what proportion of ignorants by using rumor model. They also described an impulsive control model of a rumor process to classify the spreaders [40]. Belen et al. [41] proposed a solution to the problem of a repeated rumor based on the classical Maki-Thompson rumor model, and thus they improved the Maki-Thompson model and derived a new solution for each dynamics of spreading of a rumor [42]. Similarly, Wilhelm Weber and Gürbüz [43] chose the numerical approach to study the dynamics of a rumor propagation model, and conduct detailed analysis. They then proved the effectiveness of the rumor propagation model [44]. Further, they proposed a numerical technique based on nonlinear ordinary differential equation, to solve a rumor propagation model [45].

Machine learning-based approaches usually achieve higher classification results than rule-based approaches. However, they separate the feature extraction from the decision-making process. Their performance largely depends on the feature engineering, which is often cumbersome to design.

Deep learning based approaches. As deep learning based architectures cast off the fetters of feature engineering, they usually achieve a better performance. A growing number of researchers apply deep learning technologies to sarcasm recognition as well.

As one of the first studies, Poria et al. [46] employed a pre-trained CNN for extracting sentiment, emotion and personality features. Zhang et al. [47] used a bi-directional gated recurrent neural network (RNN) to capture contextual features for sarcasm detection. Similarly, Potamias et al. [6] designed a deep framework, which consisted of the pre-trained transformer-based architecture for irony and sarcasm detection. Chaturvedi et al. [48] designed a deep CNN to extract features from texts and images, and predicted the degree of a particular emotion using a fuzzy logic classifier. Vashishtha and Susan [49] proposed an unsupervised system that was based on nine fuzzy rules to classify the posts into three sentiment classes. Chatterjee et al. [50] took the context of the utterance into consideration, and proposed a deep learning based approach. Liu et al. [51] proposed a deep neural network, called A2Text-Net, to mimic the face-to-face speech, which integrated auxiliary clues such as punctuations, part of speech (POS), emoji, etc., to improve the performance of sarcasm

detection.

Recently, contextual information has been exploited for sarcasm detection. Jaiswal [52] investigated various pre-trained language representation models and utilized the contextual information along with the utterance for SDC. Ghosh et al. [53] used LSTM and attention-based LSTM to investigate the role of conversation context in detecting sarcasm in social media discussions. For a similar purpose, Castro et al. [2] created a new dataset, MUSTARD, for multimodal sarcasm research with high-quality annotations, including both multimodal and conversational context features. In a further work [54], they argued that knowledge in sarcasm detection could also be beneficial to sentiment analysis, and designed a multi-task learning framework using a deep neural network to improve the performance of both tasks. Similarly, Chauhan et al. [5] leveraged the multi-modal sentiment and emotion information for solving the problem of multi-modal sarcasm detection in a multi-task framework. However, they neglected the contextual interaction between adjacent utterances.

To sum up, the afore-discussed three types of studies have made a good progress in sarcasm detection, and motivated our work. They are mainly focused on studying refined sarcastic rules, extracting effective features and building deep learning architectures for sarcasm recognition. However, to the best of our knowledge, they have not yet systematically taken into account the vagueness and uncertainty of human language in sarcastic expression, which we aim to tackle in this paper.

B. Fuzzy Logic based Models

To model the fuzziness in artificial intelligence problems, a large body of fuzzy logic based models has been proposed [15], [55]. For example, Alireza Goli's team has proposed a range of fuzzy models to solve cell formation problem (CFP) [14], relief vehicles problem [16], transportation route planning [55], etc [56]. Kropat and Weber [57] depicted the eco-finance networks for modeling gene-expression patterns with respect to errors and uncertainty. Erik Kropat [58] introduced time-discrete target-environment regulatory systems (TE-systems) under ellipsoidal uncertainty, and presented a mixed integer regression problem for the relaxation by means of continuous optimization. They [59] also discussed a lot of regression models for gene-environment networks under ellipsoidal uncertainty. In order to model the relations between the targets and environmental entities of the regulatory network, they proposed a fuzzy target-environment network and analyzed the vagueness of the regulatory system [60], [61]. Then, they extended the concept of fuzzy target-environment networks to fuzzy-regression models with fuzzy data sets for understanding of interconnected complex systems. [62].

To the best of our knowledge, our work is the first that brings together quantum theory and fuzzy logic for SDC. The major innovations of the work presented in this paper can be summarized as follows.

- A novel application of fuzzy logic to SDC is proposed.
- Inspired by quantum probability theory, we introduce complex numbers into the utterance representation.
- A complex-valued fuzzy network is designed, which leverages quantum probability theory and fuzzy logic

to capture both the vagueness and uncertainty in sarcastic expression.

- We verify the effectiveness of our model by applying it to the task of SDC. Empirical experimental results show that our model outperforms strong baselines.

III. Preliminaries of Quantum Theory and Fuzzy Logic

A. Quantum Theory Preliminaries

In QT, the quantum probability space is naturally encapsulated in an infinite complex Hilbert space, denoted as \mathbb{H} . The essential difference between quantum and classical probability lies in the complex nature of quantum states.

With the Dirac's notation, a pure quantum state can be represented by a ray in a Hilbert space over the complex numbers. A quantum state vector in a complex vector space, \vec{u} , can be expressed as a ket $|u\rangle$, and its transpose can be expressed as a bra $\langle u|$. In Hilbert space, a quantum system can be in multiple mutually exclusive basis states simultaneously, with a probability distribution until it is measured, called quantum superposition, namely, $|u\rangle = \sum_{i=1}^n z_i |w_i\rangle$, where the $|w_i\rangle$ are orthogonal unit vectors and the z_i are complex components. After measurement it then collapses to one of the basis states that form the superposition. Quantum superposition describes the uncertainty of a single particle. For example, if there are two basis states $|0\rangle$ and $|1\rangle$, then a superposition state would be $|u\rangle = z_\alpha |0\rangle + z_\beta |1\rangle$, where z_α and z_β are complex coefficients, satisfying $z_\alpha^2 + z_\beta^2 = 1$.

A quantum event is defined to be a subspace of Hilbert space, represented by any orthogonal projector Π . Assume $|u\rangle$ is a unit vector, i.e., $\|\vec{u}\|_2 = 1$. The projector Π is written as $|u\rangle\langle u|$. A quantum mixed state corresponds to a probabilistic mixture of pure states, which is represented by the density matrix, $\rho = \sum_\lambda \mu_\lambda |u\rangle\langle u|$. Density matrix ρ is symmetric (i.e., $\rho = \rho^T$), positive semi-definite (i.e., $\rho \geq 0$), and of trace 1. The quantum probability measure M is associated with the density matrix. The Gleason's Theorem has proven the existence of a mapping function $M(|u\rangle\langle u|) = \text{tr}(\rho|u\rangle\langle u|)$ for any $|u\rangle$. In QT, quantum measurement describes the interaction (compositing) between a quantum system and the measurement device, where the composition system can be represented by the tensor product of two systems, e.g., $M \otimes |u\rangle$.

Measurement. Measurement is a process of testing or manipulating the physical property of a system. In classical mechanics, the measurement process and measurement device are independent of the measured objects, which will not affect the measured objects. However, measurement in quantum world has an impact on the measured object, such as changing the state of the system to be measured. Quantum measurement is described by a set of measurement operators acting on the state space of the system being measured $\{M_m\}$, where m represents the possible measurement outcomes. Suppose the quantum system is in a state of $\rho|u\rangle$ before the measurement, then the probability to obtain the outcome m after the measurement is $p(m) = \langle u|M_m^\dagger M_m|u\rangle$. Moreover, the state of quantum system has changed to: $\frac{M_m|u\rangle}{\sqrt{\langle u|M_m^\dagger M_m|u\rangle}}$.

By introducing the complex number, QT could define complex probability amplitude to construct classical probability

(i.e., the square of the probability amplitude equals to the probability, providing a many-to-one relationship between probability amplitude and probability), and thus describe the uncertain events. QT provides a principled and effective mechanism to capture the intrinsic uncertainty.

B. Basic Notations and Concepts in Fuzzy Logic

The conventional Boolean logic has been applied to a wide variety of AI applications, by only permitting two truth values, i.e., true and false. It has many deficiencies since two truth values are incapable of describing complex reasoning mode of human. As an extension, fuzzy logic was invented to deal with vagueness and imprecise information by defining many values between 0 and 1, in the form of fuzzy sets [3], [63].

Fuzzy logic commonly provides a three-step process, i.e., fuzzification, logical operation and defuzzification, to map the linguistic variables onto output results. Fuzzification targets at converting the numerical input of a system x to the degree of membership in a fuzzy set A , by using membership function μ_A . The degree of membership could be any values within the interval $[0,1]$, $\mu_A(x) \in [0,1]$. There are different types of membership functions, e.g., triangular, trapezoidal, Gaussian, sigmoid, polynomial functions, etc.

Fuzzy logic has defined three basic operators, AND, OR and NOT. Assume that there are four fuzzy sets, i.e., A , B , C and D , where $C = A \cup B$, $D = A \cap B$. Then, $\mu_C(x) = \max\{\mu_A(x), \mu_B(x)\}$, $\mu_D(x) = \min\{\mu_A(x), \mu_B(x)\}$, and $\mu_{A^c}(X) = 1 - \mu_A(X)$. Defuzzification is the process of transforming the output value of a fuzzy inference system into a crisp output. There are some mostly-used algorithms, e.g., finding the center of gravity, calculating the average mean, calculating the left maximum, etc.

Building on many-valued logic, fuzzy logic aims to simulate human intelligence for automatically handling vague information, performing judgment and reasoning.

C. The Relations between QT and Fuzzy Logic

QT has a close tie with fuzzy logic, since both of them provide a mean to deal with concepts like vagueness and uncertainty. In comparison to fuzzy logic that is based on membership values, QT is defined on complex subspace identified by projectors [25]. The interaction of a projector with a density matrix produces a value which can be directly interpreted as the degree of membership. Meanwhile, the fuzzy membership function could be used to depict the relative importance of each quantum state. A few studies proved that some logic operations of projectors (e.g., junction) in QT directly corresponds to the operations in fuzzy logic under some conditions [64]. They also conducted detailed analysis of a fuzzy logic interpretation of quantum theory by demonstrating that the Schrodinger equation can be deduced from the assumptions of the fuzziness [65].

In this work, we bring QT and fuzzy logic together for modeling the intrinsic vagueness and uncertainty of human language in emotional expression and understanding.

D. How to Apply Key Notations to Our Approach

Here, we summarize the key notations in QT and fuzzy logic and explain how to apply them to our approach. (1) The quantum state $|u\rangle$ could be seen as the utterance u in conversations, while the i^{th} basis vector $|w_i\rangle$ is linked to the i^{th} basis word vector. Each utterance is thus seen as in a quantum superposition of a set of basis words, and represented as $|u\rangle = \sum_{i=1}^n z_i |w_i\rangle$. The target utterance $|u_t\rangle$ and its λ^{th} context $|c_\lambda\rangle$ are calculated in the same way (c.f. Sec. IV-C). (2) The composition system in QT is linked to the contextual interaction between the target utterance $|u_t\rangle$ and its context $|c_\lambda\rangle$, which could be calculated as the tensor product of them, e.g., $|\Psi_t^{c_\lambda}\rangle = |u_t\rangle \otimes |c_1\rangle \otimes |c_2\rangle, \dots, \otimes |c_\lambda\rangle$. (3) The final representation of the target utterance is considered as a quantum mixed state that is in a statistical mixture over multiple composition systems, which can be mathematically encapsulated in a density matrix, i.e., $\rho_t = \sum_\lambda \mu_\lambda |\Psi_t^{c_\lambda}\rangle \langle \Psi_t^{c_\lambda}|$, where μ_λ is the fuzzy membership function representing the relative importance of the λ^{th} composition system (c.f. Sec. IV-D). (4) Since we have obtained the representation of the target utterance ρ_t , a sequence of quantum fuzzy measurements $\{M_m\}$ on the representation ρ_t , for obtaining refined sarcastic features by calculating $tr(M_m \rho_t)$ (c.f. Sec. IV-E). Hence, we will benefit from the unified and principled mathematics of QT.

IV. Complex-valued Fuzzy Network

In order to capture both the vagueness and uncertainty in human language, with conversational sarcasm detection as a particular exemplar in this paper, we propose an end-to-end neural network based on QT and fuzzy logic, called complex-valued fuzzy network.

A. Problem Formulation and Overall Framework

Problem Formulation. Suppose the dataset has L samples, the γ^{th} sample X_γ could be represented as $\{X_\gamma = (C_\lambda, U_t), Y_t\}$, where C_λ , U_t , Y_t represent the λ^{th} conversational context, the target utterance and its label respectively, where $C_\lambda \in \mathbb{H}^{l_\lambda \times d_\lambda}$, $U_t \in \mathbb{H}^{l_t \times d_t}$. Here, l_λ and l_t denote the sequence length of contextual and target utterances, d_λ and d_t mean the dimensions of the contextual and textual features. $\lambda \in [1, 2, \dots, k]$, $\gamma \in [1, 2, \dots, L]$.

Now, given a conversation (including the context C_λ and target utterance U_t), how to determine the sarcasm polarity (Y_t). We formulate the problem as follows:

$$\zeta = \prod_{\lambda} p(Y_t | C_\lambda, U_t, \Theta) \quad (1)$$

where Θ represents the parameter set.

Overall Framework. The architecture of the CFN framework is shown in Fig. 2. (1) In the embedding layer, the target utterance u_t and its λ^{th} contextual utterance c_λ are embedded as complex-valued embeddings that are expressed in polar form, denoted as $|u_t\rangle$ and $|c_\lambda\rangle$. (2) In the fuzzy composition layer, the interactions between $|u_t\rangle$ and its contexts $\{|c_1\rangle, |c_2\rangle, \dots, |c_\lambda\rangle, \dots, |c_k\rangle\}$ are modeled as multiple quantum composite systems, which are given by the tensor product of individual utterance embedding, where the

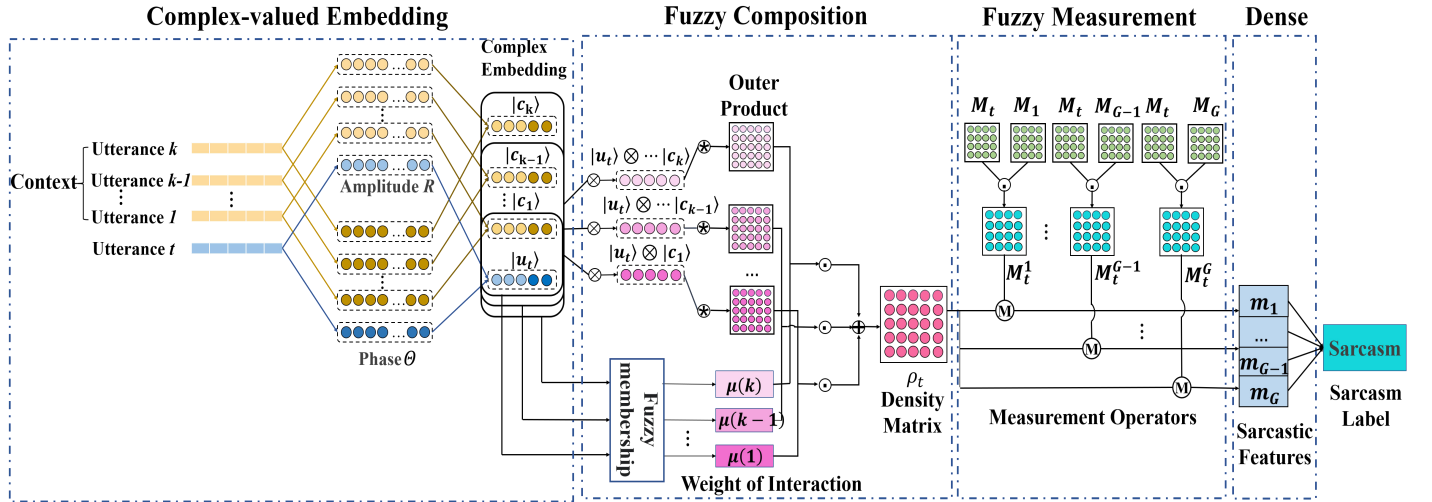


Fig. 2: The architecture of complex-valued fuzzy network. \otimes denotes the tensor product operation. \circledast denotes an outer production to a vector. \odot denotes point-wise multiplication. \oplus refers to a element-wise addition. \mathbb{M} means a quantum fuzzy measurement.

results are mathematically encapsulated in the density matrix ρ_t . (3) Sarcastic features are extracted via the concept of quantum measurement, which is a natural choice given the quantum state representation of sarcastic sentence. The fuzzy measurement layer is designed to perform a set of quantum fuzzy measurement operators to extract the sarcastic features \vec{m}_t . (4) The dense layer is employed to infer the final sarcasm polarity for the utterance.

B. Theoretical Advantages of Our CFN Framework

Before presenting the CFN framework, we provide the mathematical proofs here to show the advantages of our CFN framework in the form of three propositions.

Proposition 1 *Quantum probability is more general to capture the uncertainty in human language.*

Proof. Let $z(x) = re^{i\theta}$ be a quantum complex probability amplitude of event x . Using the definition of quantum probability, we get the classical probability of event x

$$p(x) = |z(x)|^2 = r^2,$$

that is

$$r = \sqrt{p(x)}$$

where $r \in \mathbb{R}$, $\theta \in (-\pi, \pi)$. Given $p(x)$, the complex probability amplitude will satisfy

$$z(x) = \sqrt{p(x)} \times (\cos \theta + i \sin \theta) = re^{i\theta}$$

Hence, $\exists r_1, r_2 \in \mathbb{R}, r_1 \neq r_2$ and $\forall \theta_1, \theta_2 \in (-\pi, \pi), \theta_1 \neq \theta_2$, satisfies that $z_1(x) = r_1 e^{i\theta_1}$, $z_2(x) = r_2 e^{i\theta_2}$. We obtain

$$\begin{aligned} |z_1(x)|^2 &= p(x) = |z_2(x)|^2 \\ \text{s.t. } r_1^2 &= r_2^2 \end{aligned} \quad (2)$$

We conclude that $\exists z_1, z_2 \in \mathbb{H}$ and $z_1 \neq z_2$, then $z_1 \rightarrow p \wedge z_2 \rightarrow p$.

Remark 1. For example, the probability of a word w is 0.5, i.e., $p(x=w) = \frac{1}{2}$, then the corresponding probability amplitude may be $z(x=w) = \frac{\sqrt{2}}{2} e^{i\frac{\pi}{4}}$ or $z(x=w) = -\frac{\sqrt{2}}{2} e^{i\frac{3\pi}{4}}$,

etc. There is a many-to-one relationship between complex probability amplitude and probability. The amplitude r links to the probability, while the phase θ may be associated with hidden sentiment or sarcasm orientations. An utterance thus could be represented in an amplitude-phase manner.

Proposition 2 *Quantum superposition embodies a non-linear fusion of basis states.*

Proof. Let $z_1(w_1)$ and $z_2(w_2)$ be the complex probability amplitudes of two basis words w_1, w_2 respectively, where $z_1(w_1), z_2(w_2) \in \mathbb{H}^{l_t \times d_t}$.

Let a compound term be $c \propto (w_1 w_2)$, we obtain

$$\begin{aligned} z_3(c) &= \alpha z_1(w_1) + \beta z_2(w_2) \\ \text{s.t. } \alpha^2 + \beta^2 &= 1, \\ \alpha, \beta &\in \mathbb{H} \end{aligned} \quad (3)$$

where $z_3(c) \in \mathbb{H}^{l_t \times d_t}$. Based on Proposition 1, we have

$$\begin{aligned} p(w_1) &= |z_1(w_1)|^2, \quad p(w_2) = |z_2(w_2)|^2 \\ \text{s.t. } p(w_1), p(w_2) &\in [0, 1] \end{aligned}$$

We can derive the probability of the compound term:

$$\begin{aligned} p(c) &= |z_3(c)|^2 = |\alpha z_1(w_1) + \beta z_2(w_2)|^2 \\ &= (\alpha z_1(w_1) + \beta z_2(w_2)) \cdot (\alpha z_1(w_1) + \beta z_2(w_2))^{\dagger} \\ &= \alpha z_1(w_1) \cdot (\alpha z_1(w_1))^{\dagger} + \beta z_2(w_2) \cdot (\beta z_2(w_2))^{\dagger} \\ &\quad + \alpha z_1(w_1) \cdot (\beta z_2(w_2))^{\dagger} + \beta z_2(w_2) \cdot (\alpha z_1(w_1))^{\dagger} \\ &= \alpha z_1(w_1) \cdot (\alpha z_1(w_1))^{\dagger} + \beta z_2(w_2) \cdot (\beta z_2(w_2))^{\dagger} \\ &\quad + \alpha z_1(w_1) \cdot (\beta z_2(w_2))^{\dagger} + (\alpha z_1(w_1) \cdot (\beta z_2(w_2))^{\dagger})^{\dagger} \\ &= |\alpha z_1(w_1)|^2 + |\beta z_2(w_2)|^2 + 2 \operatorname{Re}(\alpha z_1(w_1) \cdot (\beta z_2(w_2))^{\dagger}) \\ &= |\alpha z_1(w_1)|^2 + |\beta z_2(w_2)|^2 + 2 |\alpha z_1(w_1) \beta z_2(w_2)| \cos \theta \\ &= \alpha^2 p(w_1) + \beta^2 p(w_2) + 2 \alpha \beta \sqrt{p(w_1) p(w_2)} \cos \theta \end{aligned} \quad (4)$$

We use contradiction to prove the non-linearity. Assume that

$$p(x) = \alpha^2 p(w_1) + \beta^2 p(w_2) + 2\alpha\beta\sqrt{p(w_1)p(w_2)}\cos x \quad (5)$$

which is a linear function, satisfying

$$kp(x) = p(kx) \quad \forall x,$$

since this must hold for all x , it certainly must hold in the special case $x = \frac{\pi}{2}$. Let $k = 2$, then

$$\begin{aligned} & 2\alpha^2 p(w_1) + 2\beta^2 p(w_2) + 4\alpha\beta\sqrt{p(w_1)p(w_2)}\cos\frac{\pi}{2} \\ &= \\ & \alpha^2 p(w_1) + \beta^2 p(w_2) + 2\alpha\beta\sqrt{p(w_1)p(w_2)}\cos\frac{2\pi}{2} \end{aligned} \quad (6)$$

which leads to the ridiculous conclusion that

$$\alpha^2 p(w_1) + \beta^2 p(w_2) = -2\alpha\beta\sqrt{p(w_1)p(w_2)} \quad (7)$$

Our assumption that Eq. 5 is linear is false.

Remark 2. Hence, the probability of the compound term is the non-linear superposition of the probabilities of the basis words, with an interference term determined by the relative phase θ . This provides a higher level of abstraction.

Proposition 3 *Quantum composition system describes the correlations between parts and whole.*

Proof. Let u_i and u_j represent two adjacent utterances, we obtain

$$\begin{aligned} |u_i\rangle &= \alpha_1|w_1\rangle + \beta_1|w_2\rangle \\ |u_j\rangle &= \alpha_2|w_1\rangle + \beta_2|w_2\rangle \\ \text{s.t. } \alpha_1^2 + \beta_1^2 &= 1, \\ \alpha_2^2 + \beta_2^2 &= 1, \\ \alpha_1, \alpha_2, \beta_1, \beta_2 &\in \mathbb{H} \end{aligned} \quad (8)$$

State space of a composite system \mathcal{H}_{u_i, u_j} consisting of two utterances u_i and u_j is written as a tensor product of the individual state spaces $|u_i\rangle$ and $|u_j\rangle$:

$$\begin{aligned} \mathcal{H}_{u_i, u_j} &= |u_i\rangle \otimes |u_j\rangle \\ &= (\alpha_1|w_1\rangle + \beta_1|w_2\rangle) \otimes (\alpha_2|w_1\rangle + \beta_2|w_2\rangle) \\ &= \alpha_1|w_1\rangle \otimes (\alpha_2|w_1\rangle + \beta_2|w_2\rangle) \\ &\quad + \beta_1|w_2\rangle \otimes (\alpha_2|w_1\rangle + \beta_2|w_2\rangle) \\ &= \alpha_1\alpha_2|w_1w_1\rangle + \alpha_1\beta_2|w_1w_2\rangle \\ &\quad + \beta_1\alpha_2|w_2w_1\rangle + \beta_1\beta_2|w_2w_2\rangle \end{aligned} \quad (9)$$

Let $|w_1\rangle = (x_1, x_2)^T$, $|w_2\rangle = (y_1, y_2)^T$, then

$$\begin{aligned} \mathcal{H}_{u_i, u_j} &= \alpha_1\alpha_2 \begin{bmatrix} x_1^2 & x_1x_2 \\ x_2x_1 & x_2^2 \end{bmatrix} + \alpha_1\beta_2 \begin{bmatrix} x_1y_1 & x_1y_2 \\ x_2y_1 & x_2y_2 \end{bmatrix} \\ &\quad + \beta_1\alpha_2 \begin{bmatrix} y_1x_1 & y_1x_2 \\ y_2x_1 & y_2x_2 \end{bmatrix} + \beta_1\beta_2 \begin{bmatrix} y_1^2 & y_1y_2 \\ y_2y_1 & y_2^2 \end{bmatrix} \end{aligned} \quad (10)$$

where \mathcal{H}_{u_i, u_j} is controlled by the basis words.

Remark 3. Eq. 10 proves that the composition system consisting of utterances embodies the correlations between words, which inspires us to model the contextuality by a “global to local” way. The details of the CFN framework will be given in the next subsections.

C. Complex-valued Utterance Embedding

In QT, quantum probability could be seen as the mathematical product of generalizing classical probability to the complex number field. Complex numbers are introduced to describe uncertain quantum state and behavior by defining the concept of quantum probability amplitude, where the modulus squared of the quantum probability amplitude represents a classical probability, providing a many-to-one relationship between probability amplitude and probability. Complex probability amplitude possesses outstanding ability to model the uncertainty.

Motivated by Wang and Li’s work [66], we seek inspirations from quantum probability, and design a complex-valued utterance embedding layer. As word is the basic semantic unit of human language, we regard each word w_j in conversations as the basis state $|w_j\rangle$, and assume that $\{|w_1\rangle, |w_2\rangle, \dots, |w_n\rangle\}$ builds the orthogonal basis of dialogic Hilbert space \mathbb{H}_d . The basis state $|w_j\rangle$ is mapped into dialogic Hilbert space \mathbb{H}_d using one hot encoding, which consists of zero in all cells with the exception of a single one in a cell used uniquely to identify the word, i.e., $|w_j\rangle = \begin{pmatrix} 0, 0, \dots, 0, 1, 0, \dots, 0 \end{pmatrix}_T$.

Then, an utterance could be seen as a collection of words. In QT, superposition is a fundamental concept, which has described the inherent uncertainty in the state of a microscopic particle (which, in this paper, refers to human sarcastic attitude). In order to capture the uncertainty in human language, we regard the target utterance u_t as a quantum superposition of a set of basis words $\{|w_1\rangle, |w_2\rangle, \dots, |w_n\rangle\}$, which can be formulated as:

$$\begin{aligned} |u_t\rangle &= \sum_{j=1}^n z_j |w_j\rangle \\ z_j &= r_j e^{i\theta_j} \end{aligned} \quad (11)$$

where n is the number of words in the utterance. z_j is a complex probability amplitude that is expressed in polar form, sufficing $\sum z_j^2 = 1$. i is the imaginary number satisfying the equation $i^2 = -1$. r_j represents the absolute value or modulus of the complex vector, termed amplitude. $\theta_j \in (-\pi, \pi)$ is the argument (phase) of z_j , referring to the angle of the ray with the positive real axis.

Let the amplitude \mathbf{R} represent the length of the ray while the phase $\mathbf{\Theta}$ denotes its direction. In this paper, we associate both of them with specific linguistic meanings. The amplitude is analogous to the semantic knowledge. As for the phase, since sarcasm is closely related to sentiment, it is linked to the sentiment orientation of the utterance. Now, the target utterance has been represented as a complex-valued vector, i.e., $|u_t\rangle = (r_1 e^{i\theta_1}, r_2 e^{i\theta_2}, \dots, r_n e^{i\theta_n})^T$. The embedding of λ^{th} contextual utterance, i.e., $|c_\lambda\rangle$, can be calculated in the same way.

Eq. 11 is proposed to map each utterance into an embedding and learn its complex-valued vector representation. This complex-valued representation could capture the uncertainty in human language via the concept of quantum superposition and complex probability amplitude. This representation can be used to model the interaction between the target utterance and its contexts as a basic form (c.f. Eq. 13).

Moreover, this representation uses both the amplitude and phase to determine the complex probability amplitude z_j , releasing strong ability to model the probability. It also allows for a non-linear fusion of amplitudes and phases in its mathematical form. Suppose the components $r_j^1 e^{i\theta_j^1}$ and $r_j^2 e^{i\theta_j^2}$ represent the j^{th} dimension of two utterances u_1 and u_2 respectively. The component $r_j^3 e^{i\theta_j^3}$ of the combination of u_1 and u_2 could be computed as:

$$\begin{aligned} r_j^3 e^{i\theta_j^3} &= r_j^1 e^{i\theta_j^1} + r_j^2 e^{i\theta_j^2} \\ &= \sqrt{|r_j^1|^2 + |r_j^2|^2 + 2r_j^1 r_j^2 \cos(\theta_j^1 - \theta_j^2)} \\ &\quad \times e^{i \arctan\left(\frac{r_j^1 \sin(\theta_j^1) + r_j^2 \sin(\theta_j^2)}{r_j^1 \cos(\theta_j^1) + r_j^2 \cos(\theta_j^2)}\right)} \end{aligned} \quad (12)$$

Eq. 12 provides a non-linear fusion way to describe the compound term or phrase.

D. Learning Contextual Interaction with The Fuzzy Composition Layer

In QT, ideal quantum measurement describes the interaction (coupling) between a quantum system, the measurement device and their surrounding environments (e.g., other neighbor systems). However, we argue that the measurement device and the environments may not participate equally in the interaction with the quantum system in realistic world, i.e., their degrees of involvement in the interaction are not equal. This interaction is analogous to the interaction among utterances in a conversation, while different context utterances, e.g., in the previous turns of a conversation, express different intensities of human interaction.

Inspired by this observation, we treat the target utterance u_t as a quantum system to be measured, its contextual utterances $C = \{c_1, c_2, \dots, c_k\}$ as the surrounding environments. The measurement device is characterized as a set of measurement operators, which will be detailed in next subsection (c.f. Sec. IV-E). Based on fuzzy logic, we thus design a fuzzy composition layer to capture contextual interaction among utterances.

Suppose that a target utterance u_t is represented by its complex-valued embedding $|u_t\rangle$, and its contexts are represented by their embeddings $\{|c_\lambda\rangle\}_{\lambda=1}^k = \{|c_1\rangle, |c_2\rangle, \dots, |c_k\rangle\}$ respectively. The interaction between the target utterance u_t and its λ^{th} context c_λ constructs a composite system $|\Psi_t^{c_\lambda}\rangle$, which is given by the tensor product of individual utterance embedding. In this work, the intensity of human interaction is dependent on the distance of context utterances directly. We aim to learn both long and short range contextual interactions, by constructing multiple composite systems with a variable number of contexts. The λ^{th} composite system could be computed as:

$$\begin{aligned} |\Psi_t^{c_\lambda}\rangle &= |u_t\rangle \otimes \{|c_q\rangle\}_{q=1}^\lambda \\ &= |u_t\rangle \otimes |c_1\rangle \otimes |c_2\rangle \otimes \dots \otimes |c_\lambda\rangle \end{aligned} \quad (13)$$

where $\lambda \in [1, k]$. According to Eq. 13, we have built k composite systems for k context utterances, i.e., $\Psi_k = \{|\Psi_t^{c_1}\rangle, |\Psi_t^{c_2}\rangle, \dots, |\Psi_t^{c_k}\rangle\}$. The role of Eq. 13 is to learn the interactions between the target utterance and its contexts, by introducing the operation of tensor product, where the result is represented as the composite system.

Then, the final representation of the target utterance is in a statistical mixture over multiple composite systems Ψ_k , which can be mathematically encapsulated in a density matrix. Suppose the λ^{th} composite system $|\Psi_t^{c_\lambda}\rangle = (\Psi_{t,1}^{c_\lambda}, \Psi_{t,2}^{c_\lambda}, \dots, \Psi_{t,d}^{c_\lambda})$, the density matrix of utterance u_t can be represented as:

$$\begin{aligned} \rho_t &= \sum_{\lambda=1}^k \mu_\lambda |\Psi_t^{c_\lambda}\rangle \langle \Psi_t^{c_\lambda}| \\ &= \begin{bmatrix} \sum_{\lambda} \mu_\lambda (\Psi_{t,1}^{c_\lambda})^2 & \sum_{\lambda} \mu_\lambda \Psi_{t,1}^{c_\lambda} \Psi_{t,2}^{c_\lambda} & \dots & \sum_{\lambda} \mu_\lambda \Psi_{t,1}^{c_\lambda} \Psi_{t,d}^{c_\lambda} \\ \sum_{\lambda} \mu_\lambda \Psi_{t,2}^{c_\lambda} \Psi_{t,1}^{c_\lambda} & \sum_{\lambda} \mu_\lambda (\Psi_{t,2}^{c_\lambda})^2 & \dots & \sum_{\lambda} \mu_\lambda \Psi_{t,2}^{c_\lambda} \Psi_{t,d}^{c_\lambda} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{\lambda} \mu_\lambda \Psi_{t,d}^{c_\lambda} \Psi_{t,1}^{c_\lambda} & \sum_{\lambda} \mu_\lambda \Psi_{t,d}^{c_\lambda} \Psi_{t,2}^{c_\lambda} & \dots & \sum_{\lambda} \mu_\lambda (\Psi_{t,d}^{c_\lambda})^2 \end{bmatrix} \end{aligned} \quad (14)$$

where μ_λ is the fuzzy membership function representing the relative importance of the λ^{th} composite system. We argue that there is no firm boundary among the degrees of the participation in interaction, and thus adopt a Sigmoidal fuzzy membership function to model the fuzziness. μ_λ is decided by $\mu(\lambda : a, c) = \frac{1}{1 + e^{-a(\lambda - c)}}$, where λ is the index of context utterance, a and c are trainable parameters learned during training.

Eq. 14 is used to obtain the matrix representation of the target utterance, and has shown each element of matrix. The density matrix ρ_t has encoded all the information and properties of the target utterance u_t , e.g., the semantic dependencies, the probabilistic distribution information. The target utterance $|u_t\rangle$ and its contexts $\{|c_\lambda\rangle\}_{\lambda=1}^k$ are unified into a density matrix, for capturing both the uncertainty and vagueness simultaneously. The density matrix ρ_t is thus fed into the quantum fuzzy measurement layer to extract more refined sarcastic features.

E. Quantum Fuzzy Measurement

In QT, measurement is a process of testing or manipulating the physical property of a quantum system, e.g., the movement locus of a particle. The outcome of quantum measurement is associated with a set of numerical values. Before performing quantum measurement, there is uncertainty in the system in that it takes all possible measurement values simultaneously. After the measurement, the system collapses into one of these basis values, and the uncertainty on the system state is hence removed.

Further, the information and property of a system (e.g., an utterance's sarcastic features) could be depicted by the probability distribution of the measurement outcomes, which are acquired by performing a finite set of quantum measurements on the system [67]. As an extension of standard quantum measurement, quantum fuzzy measurement argues that the measurement device does not participate completely in interaction with the quantum system, meaning that the measurement device is not necessarily orthogonal to the quantum system, and could be set freely. Motivated by this, we perform a sequence of quantum fuzzy measurements with different measuring angles on the composite system, for obtaining rich sarcastic features for the target utterance $\vec{m}_t = (m_1, m_2, \dots, m_G)$.

We define a set of fuzzy measurement operators $\{M_t^\delta\}_{\delta=1}^G$, each of which is constructed by two sub-measurement M_t and $\{M_\delta\}_{\delta=1}^G$. M_t represents a trainable measurement operator

(i.e., matrix) for the target utterance u_t , while $\{M_\delta\}_{\delta=1}^G$ represents another trainable measurement operator. The δ^{th} fuzzy measurement operator M_t^δ is a $d \times d$ matrix, satisfying the completeness condition that:

$$\sum_{\delta=1}^G (M_t^\delta)^\dagger M_t^\delta = 1 \quad (15)$$

In QT, based on the density matrix of utterance u_t , the probabilistic expected value of quantum fuzzy measurement result, denoted as $\langle m_\delta \rangle$, is calculated as:

$$\begin{aligned} \langle m_\delta \rangle &= \sum_{\lambda} \mu_{\lambda} \langle \Psi_t^{c_{\lambda}} | M_t^\delta | \Psi_t^{c_{\lambda}} \rangle \\ &= \sum_{\lambda} \mu_{\lambda} \text{tr}(M_t^\delta | \Psi_t^{c_{\lambda}} \rangle \langle \Psi_t^{c_{\lambda}} |) \\ &= \sum_{\lambda} \text{tr}(\mu_{\lambda} M_t^\delta | \Psi_t^{c_{\lambda}} \rangle \langle \Psi_t^{c_{\lambda}} |) \\ &= \text{tr}\left(\sum_{\lambda} \mu_{\lambda} M_t^\delta | \Psi_t^{c_{\lambda}} \rangle \langle \Psi_t^{c_{\lambda}} | \right) \\ &= \text{tr}(M_t^\delta \rho_t) \end{aligned} \quad (16)$$

Eq. 16 computes the expected value of the δ^{th} measurement result, and leads to the following calculation about the probability distribution over the measurement outcomes. We regard this probability distribution as the feature vector (i.e., \vec{m}_t) for the target utterance. The procedure is written as:

$$\begin{aligned} m_\delta &= \text{tr}(M_t^\delta (M_t^\delta)^\dagger \rho_t) \\ &= \text{tr}\left((M_t^\delta)^\dagger M_t^\delta \sum_{\lambda} \mu_{\lambda} | \Psi_t^{c_{\lambda}} \rangle \langle \Psi_t^{c_{\lambda}} | \right) \\ &= \sum_{\lambda} \mu_{\lambda} \text{tr}\left((M_t^\delta)^\dagger M_t^\delta | \Psi_t^{c_{\lambda}} \rangle \langle \Psi_t^{c_{\lambda}} | \right) \\ &= \sum_{\lambda} \langle \Psi_t^{c_{\lambda}} | (M_t^\delta)^\dagger M_t^\delta | \Psi_t^{c_{\lambda}} \rangle \end{aligned} \quad (17)$$

where $M_t^\delta = M_t M_\delta$, $\delta \in [1, 2, \dots, G]$. Eq. 17 is computed to obtain the probabilistic value of the δ^{th} fuzzy measurement outcome, which is treated as the δ^{th} eigenvector of \vec{m}_t . Finally, we have obtained a probabilistic feature vector $\vec{m}_t = (m_1, m_2, \dots, m_\delta, \dots, m_G)$.

F. Dense Layer

The probabilistic features \vec{m}_t is passed to a fully connected layer whose output is used as the final utterance representation $\vec{x}_t = (x_1, x_2, \dots, x_\kappa)$. \vec{x}_t is fed into the softmax function for sarcasm classification.

G. Classification

We put \vec{x}_t into a *softmax* function to yield the sarcasm label $Y \in \{y_{sar}, y_{non}\}$. That is,

$$\hat{y} = \text{softmax}(W_y \vec{x}_t + b_y) \quad (18)$$

where W_y and b_y are the weight and bias.

Model training. In our framework, cross entropy with L_2 regularization is used as the loss function, which is defined as:

$$J = -\frac{1}{N} \sum_t \sum_{\xi} y_t^{\xi} \log \hat{y}_t^{\xi} + \tau_r \|\phi\|^2 \quad (19)$$

where y_t denotes the ground truth, \hat{y}_t is the predicted sarcasm distribution. t is the index of utterance, ξ is the index of class, and L is the total number of utterances. τ_r is the coefficient for L_2 regularization. We use the back propagation method to compute the gradients and update all the parameters. In order to avoid overfitting, we use dropout strategy.

V. Experiments

In this section, we validate the theoretical advantages of the QPM framework from an experimental viewpoint.

A. Experimental Settings

Our main research questions are:

- (1) Is the complex-valued fuzzy representation effective for modeling human language?
- (2) Does modeling of conversation context help in conversational sarcasm detection?
- (3) Which component of CFN plays a key role on the performance?

To answer (1), we compare the performance of the CFN model with a number of baselines, and analyze the importance of the complex-valued fuzzy representation. To answer (2), we also perform all the experiments for the target utterances without context information, and make a detailed contrast. To answer (3), we conduct an ablation test by removing one component at a time and evaluate their impacts.

Datasets. Given that SDC is a new area, the benchmark datasets are relatively limited. In this work, we perform experiments on the MUSTARD² [2] and the 2020 sarcasm detection Reddit track³ [68] datasets. MUSTARD comprises 690 videos from several sources, e.g., Big Bang Theory, Friends, etc. The utterance in each dialogue is annotated with sarcastic or non-sarcastic label. Each utterance and its context consists of three modalities: video, audio, and text. Also, all the utterances are accompanied by their speaker identifiers. In this work, we focus on conversational sarcasm detection only from the textual information. Multimodal sarcasm detection is left as future work.

The 2020 sarcasm detection Reddit track (the Reddit track for short) is a subset of the Reddit Corpus, which consists of sarcastic responses and their contexts (quotes to which the posts are replies to). It contains 3,100 posts per class (sarcastic and not-sarcastic, for a total of 6200 posts) and more than 18,000 contextual utterances. Table I shows the detailed statistics for these two datasets.

Pre-processing. Apart from the standard preprocessing steps such as lowercasing the letters, removal of emojis, and correcting spelling mistakes, we remove the stop words using a standard English stopword list. Note that we do not filter out the punctuations, since they tend to carry subjective information.

²<https://github.com/soujanyaaporia/MUSTARD>

³<https://github.com/EducationalTestingService/sarcasm/releases>

Dataset	No. of Utterance		No. of Contexts
	Sarcastic	Non-sarcastic	
MUStARD	345	345	2261
the Reddit track	3100	3100	18618

TABLE I: Statistics of the datasets.

Evaluation metrics. We adopt the **precision, recall, micro F1 score, accuracy, sensitivity and specificity** as the evaluation metrics to evaluate the classification performance. We run the experiments using five-fold cross-validation on all the comparative models.

Hyperparameters setting. In this work, the amplitudes are initialized with BERT semantic representation. The phases are set to the sentiment orientations of utterances using BERT. The fuzzy measurements are randomly initialized with an unit matrix. All weight matrices are given their initial values by sampling from a uniform distribution $U(-0.1, 0.1)$, and all biases are set to zeros. The coefficient of $L2$ normalization in the objective function is set to 10^{-5} , and the dropout rate is set to 0.5.

We search for the best performance from a parameter pool, which contains a learning rate in $\{0.001, 0.005, 0.01\}$, the batch size in $\{32, 64\}$ and the number of fuzzy measurements in $\{100, 300, 500, 1000, 1500\}$.

B. Comparative Models

In order for a comprehensive evaluation of the CFN framework, we include a range of baselines for comparison. They are listed as follows.

SVM+BERT [69] represents the textual utterances using BERT vectors and standard hyperparameter settings for sarcasm detection. We set the kernel function to “RBF”. We also attempt to concatenate the contextual features.

CNN [70] contains two convolutional layers and a fully connected layer, which is trained on top of word embeddings for sarcasm classification in conversation.

BiGRU [71] leverages a bidirectional GRU network for learning utterance representations and then uses a classification layer to make prediction of sarcasm.

MHA-BiLSTM [36] extracts the most significant features and builds a multi-head attention-based bidirectional long-short memory network to detect sarcastic utterances.

Contextual LSTM [72] first uses the CNN to extract context-independent textual features and thus feeds them into the LSTM network to obtain context-sensitive feature representations and sarcasm labels for each utterance.

RCNN-RoBERTa [6] utilizes pre-trained RoBERTa vectors combined with a RCNN in order to capture contextual information for sarcasm classification.

C-Net [73] uses BERT and simple exponential smoothing to represent each utterance in conversation thread, and takes contextual information of a sentence in a sequential manner to classify it as sarcastic or non-sarcastic.

MTL framework [5] proposes two attention mechanisms, i.e., inter- and intra-segment inter-modal attentions, to learn the relationship between the different segments and the relationship within the same segment. Representations from both the

attentions are concatenated for sentiment and sarcasm analysis. In this paper, we only use bimodal information (e.g., texts and images).

C. Results and Analysis on MUStARD

The experimental results are summarized in Table II. Since the sample size of MUStARD is relatively small, we will pay more attention to F1 score and sensitivity here. We could notice that those approaches including SVM+BERT, BiGRU, MHA-BiLSTM, Contextual LSTM, RCNN-RoBERTa, C-Net, MTL and CFN are all superior than CNN. Because all of them deal with the contextual information, which highlights the importance of modeling the contextual information in SDC. Among all baseline approaches, SVM+BERT achieves a higher F1 score than three LSTM variants (i.e., BiGRU, MHA-BiLSTM and Contextual LSTM). The reason is that the pre-trained textual features provided by BERT have stronger discrimination ability. By integrating the contextual information into the final features, SVM+BERT (+context) have obtained a slight improvement. Among three LSTM variants, BiGRU and Contextual LSTM get comparable F1 score, while MHA-BiLSTM performs worst. But MHA-BiLSTM has higher recall and sensitivity scores, which shows that MHA-BiLSTM could classify more true samples. Contextual LSTM gets the highest specificity scores among all baselines, which implies that Contextual LSTM tends to identify negative (e.g., non-sarcastic) samples. This shows that: (1) preserving the sequential order of utterances is insufficient to effectively model the conversation context; (2) the attention mechanism does not show obvious effects on improving performance on this dataset.

Through presenting improved modifications for training BERT models, RCNN-RoBERTa outperforms SVM+BERT in term of F1. But it achieves lower sensitivity and specificity scores than SVM+BERT. The possible reason is that RCNN-RoBERTa is less sensitive to sarcastic utterances than SVM classifier. As a state-of-the-art conversational sarcasm detection baselines, C-Net leverages inter-speaker dependency of the speakers to model conversational context. It performs very well in SDC. MTL framework achieves the best classification results among all baselines. Compared with C-Net, the sensitivity, f1 and accuracy results increase by 1.9%, 1.8% and 1.7%, respectively. The reason is that it designs a multi-task learning framework for multi-modal sarcasm, sentiment analysis. It leverage the utility of sentiment and emotion of the speaker to predict sarcasm, based on textual and visual modalities. Finally, CFN outperforms the state-of-the-art method MTL by margins of 8.0%, 8.0% and 8.0%. CFN achieves best scores over almost all evaluation metrics, showing that our model could not only distinguish sarcastic utterances from all testing samples, but also retrieve more sarcastic utterances from all sarcastic samples. CFN introduces complex-valued utterance embedding, where the amplitude is analogous to the semantic knowledge, while the phase is linked to the subjective information. CFN brings the semantic and subjective information together by using this representation, which can improve the description of the sarcastic information. This explains its higher sensitivity when compared to previous

Dataset	Method	Evaluation metrics					
		Precision	Recall	F1	Accuracy	Sensitivity	Specificity
MUS _T ARD	SVM+BERT	0.651	0.646	0.647	0.646	0.646	0.643
	SVM+BERT (+context)	0.655	0.651	0.650	0.651	0.651	0.650
	CNN	0.579	0.569	0.571	0.589	0.569	0.533
	BiGRU	0.631	0.643	0.624	0.656	0.643	0.579
	MHA-BiLSTM	0.673	0.652	0.614	0.668	0.652	0.644
	Contextual LSTM	0.642	0.627	0.620	0.662	0.627	0.740
	RCNN-RoBERTa	0.687	0.643	0.651	0.671	0.643	0.511
	C-Net	0.679	0.683	0.681	0.683	0.683	0.678
	MTL	0.694	0.694	0.693	0.695	0.694	0.690
	CFN	0.753 (+8.0%)	0.755 (+8.0%)	0.754 (+8.0%)	0.754 (+8.0%)	0.755 (+8.0%)	0.701 (+1.4%)

TABLE II: Performance of all baselines on MUS_TARD. The best performing system is indicated in bold. Numbers in parentheses indicate relative improvement over the MTL framework.

models. CFN obtains the second highest specificity score among all methods, showing that discovering non-sarcastic utterances is the weak point. We attribute the main improvements to both the complex valued embedding and the fuzzy composition mechanism, which ensures that CFN could preserve semantic and sentiment information, capture the figurative language’s vagueness and model the previous speaker’s influence.

D. Results and Analysis on the Reddit track

Table III presets the performance comparison of CFN with the baselines on the Reddit track. Compared with MUS_TARD, the Reddit track involves longer utterances and only contains textual modality and sarcasm label.

From Table III, we can first notice the poor performance of BiGRU and Contextual LSTM. They get the worst and the second worst sensitivity and F1 scores. This phenomenon may be because the long range of contextual utterances makes it difficult for them to record useful information into the memory unit. Contextual LSTM gets a higher specificity scores over BiGRU, showing that Contextual LSTM tends to identify negative (e.g., non-sarcastic) samples. By introducing the attention mechanism to assign greater weights to important contexts, MHA-BiLSTM outperforms BiGRU and Contextual LSTM. Compared with MHA-BiLSTM, CNN gets comparable F1 score, cause that CNN may find relatively good local optima. But CNN achieves lower sensitivity result than MHA-BiLSTM. The reason is that MHA-BiLSTM may focus on the sarcastic information by using the attention mechanism, which recognizes more sarcastic utterances and leads to a better sensitivity than CNN. Similarly, SVM+BERT overcomes the aforementioned baselines in terms of almost all evaluation metrics. This shows the fact again that the pre-trained textual features provided by BERT have great discrimination ability. The reason is that BERT’s word vector output encodes syntax trees in the word embeddings it outputs for a sentence, and encodes syntactic and semantic features in word vectors in complementary subspaces. Further, SVM+BERT (+context) performs better than SVM+BERT, by considering the contextual features, which implies that the important role of contexts.

RCNN-RoBERTa achieves the best classification performance among all approaches, whose sensitivity, F1 and accuracy scores exceed 70%. Because RCNN-RoBERTa is

trained with dynamic masking, sentences without next sentence prediction loss, larger batches, and a larger byte-level BSE. It is based on a hybrid deep neural architecture that utilizes pre-trained transformer models and feeds the hidden representations of the transformer into a recurrent convolutional neural network, which can capture unbiased recurrent informative relationships within text. It is more suitable for applying to large size dataset. This can explain its main achievement on f1 and sensitivity results. C-Net is superior than SVM+BERT (+context) but it is weaker than RCNN-RoBERTa. This may indicate RoBERTa performs better than BERT (base-uncased) on this track. Through combining both textual and visual modalities, MTL outperforms C-Net but it is still not suppress RCNN-RoBERTa. MTL does not perform well in uncovering non-sarcastic utterances. Because MTL is designed for multi-task learning problems, which has considered mutual assistance between the two tasks. The generalization ability of MTL is insufficient to deal with various single task learning problems. RCNN-RoBERTa outperforms MTL by margins of 6.0%. From an overall perspective, it seems that the baselines that have used the transformer-architecture perform better than those baselines that have used RNN-architecture. Transformer models are essentially attention based models. They regard the sentence as a whole unlike RNNs where the sentence is processed sequentially, i.e., one word per time step. Transformer models can capture long term dependencies very naturally given the attention mechanism. These are the benefits of Transformers over RNNs, showing the effectiveness of pre-trained transformer model. The proposed CFN obtains comparable specificity, sensitivity, F1 and accuracy scores against RCNN-RoBERTa, which achieves the second best classification performance. CFN brings the semantic and subjective information together by using this representation, which can improve the description of the sarcastic information. It encodes the contextual interactions between adjacent utterances into a density matrix. This explains its higher sensitivity when compared to other models. Meanwhile, according to our preliminary analysis, the failure to adopt the attention attention is one possible reason why CFN obtains lower sensitivity and F1 results. But CFN involves fewer parameters and simpler structures than RCNN-RoBERTa. We argue that CFN strikes a balance between effectiveness and efficiency.

Dataset	Method	Evaluation metrics					
		Precision	Recall	F1	Accuracy	Sensitivity	Specificity
Reddit track	SVM+BERT	0.640	0.642	0.638	0.640	0.642	0.647
	SVM+BERT (+context)	0.649	0.647	0.643	0.651	0.647	0.659
	CNN	0.631	0.623	0.629	0.623	0.623	0.611
	BiGRU	0.600	0.599	0.600	0.600	0.599	0.561
	MHA-BiLSTM	0.629	0.632	0.626	0.636	0.632	0.610
	Contextual LSTM	0.613	0.609	0.610	0.615	0.609	0.658
	RCNN-RoBERTa	0.708	0.708	0.708	0.708	0.708	0.682
	C-Net	0.663	0.664	0.663	0.663	0.664	0.657
	MTL	0.667	0.668	0.667	0.668	0.668	0.625
	CFN	0.680	0.680	0.680	0.680	0.680	0.664
		(+1.5%)	(+1.5%)	(+1.5%)	(+1.5%)	(+1.5%)	(+6.2%)

TABLE III: Performance of all baselines on the Reddit track. The best performing system is indicated in bold. Numbers in parentheses indicate relative improvement over the MTL framework.

E. Effect of Context Range

Dataset	Context range	Metrics	
		F1	Accuracy
MUSARD	Zero	0.640	0.641
	One	0.725	0.725
	Two	0.699	0.699
	All	0.754	0.754
Reddit	Zero	0.596	0.596
	One	0.641	0.641
	Two	0.617	0.617
	All	0.680	0.680

TABLE IV: Effect of context range.

We report results for CFN in Tables IV with different context scopes. Zero context means that we only use the complex-valued vector to represent the target utterance, ignoring the contextual interaction. One context utterance denotes that we use one history utterance before the target utterance to construct the density matrix. Two contexts mean that we use the previous two history utterances to construct the density matrix. From Table IV, we could observe that CFN with zero context range expectedly performs worst on both MUSARD and the Reddit track. Because it does not consider any contextual information. This shows the important role of conversation contexts. Among three methods, CFN with one context range gets the best F1 scores of 72.5% and 64.1% on two datasets, which implies that modeling the previous one history utterance is of great help to improving performance. However, its performance is weaker than the performance of CFN with all context utterances. It shows that taking all conversation contexts into consideration may be the best way to reach optimal performance.

F. Ablation Study

In this subsection, we design a series of models to further demonstrate the effectiveness of our proposed CFN model: (1) *CFN-Real* that does not consider the complex embedding, i.e., replacing utterance embeddings with their real counterparts; (2) *CFN-Speaker Independent* model without modeling the contextual interaction, which only uses the complex vector to represent the target utterance; (3) *CFN-CNN*, which replaces the quantum fuzzy measurement with the convolutional and pooling layers, to explore the significance of quantum measurement.

Dataset	Models	Metrics	
		F1	Accuracy
MUSARD	CFN-Real	0.644	0.645
	CFN-Speaker Independent	0.640	0.641
	CFN-CNN	0.740	0.740
	CFN	0.754	0.754
Reddit	CFN-Real	0.627	0.626
	CFN-Speaker Independent	0.596	0.596
	CFN-CNN	0.675	0.675
	CFN	0.680	0.680

TABLE V: Ablated CFN for both MUSARD and Reddit datasets.

In Table V, we observe that *CFN-CNN* and CFN get the best performance among these models. *CFN-Real* performs worse but still outperforms BiGRU and MHA-BiLSTM. This is because the fuzzy composition mechanism could incorporate rich interactive information into an unified density matrix. *CFN-Speaker Independent* under-performs *CFN-Real*, showing that the effectiveness of complex-valued representation and the key role of conversation contexts. Because complex-valued representation might encode non-linear semantic and sentiment compositionality. Finally, *CFN-CNN*, which replaces the quantum fuzzy measurement with the convolutional and pooling layers, obtains better results than *CFN-Real* and *CFN-Speaker Independent*, showing that the complex-valued embedding and fuzzy composition mechanism play the most important role in improving classification performance. *CFN-CNN* gets comparable results against CFN. However, CFN is theoretically more principled. Hence, these experimental results verify the effectiveness of CFN.

G. Sensitivity Study

In order to understand the relationships between inputs and outputs and study how the uncertainty in the output of the CFN model, we conduct the sensitivity analysis. In order to explore how much each input parameter is contributing to the output uncertainty, we choose one at a time (OAT) strategy, which is a common and effective sensitivity analysis method. We modify one input parameter and keep others at their original values. Then we repeat for each of the other parameters in the same way, to see the changes in the output.

No.	Utterances	Sarcasm	
		Actual	CFN
1	<i>Good idea, sit with her. Hold her, comfort her. And if the moment feels right, see if you can cop a feel.</i>	S	NS
2	<i>Just the latest copy of Applied Particle Physics quarterly.</i>	S	NS
3	<i>but but...she did them herself, so that's like, original and so edgy!</i>	S	NS
4	<i>I'm sorry, I am not going back to the Renaissance fair.</i>	NS	S
5	<i>Nah, they're too afraid of a lawsuit and bad press.</i>	NS	S

TABLE VI: Few error cases for CFN.

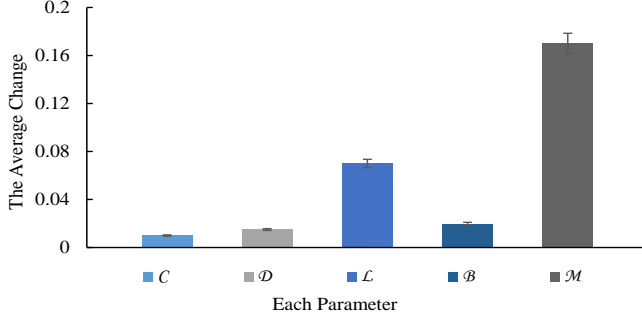


Fig. 3: Sensitivity study.

Due to the large amounts of parameters in deep neural networks, we mainly discuss five key parameters, which are the coefficient of $L2$ normalization (denoted as C), the dropout rate (denoted as D), the learning rate (denoted as L), batch size (denoted as B) and the number of measurement operators (denoted as M). Fig.3 shows the average change in the output value of the CFN framework.

From Fig.3, we notice that the CFN framework is the most sensitive to the number of measurement operators. The average change in the output value has reached about 17%. The second is learning rate L , where the average change is 7%. Then, the coefficient of $L2$ normalization, batch size and the dropout rate may lead to a slight change. These results indicate that we may pay more attention to tuning the learning rate L and the number of measurement operators M . There are two possible explanations: (1) the number of measurement operators determines the dimensionality of sarcastic feature space, which directly influences captured information. Intuitively, the more the number of measurement operators, the greater the performance of the CFN model; (2) the learning rate has been considered as the most important hyperparameter in neural network. A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck.

H. Error Analysis

We have analyzed the prediction sarcasm labels and found that misclassification often happens in the situation where the speaker uses a few positive words to express his/her sarcastic attitude. In this case, the speaker is happy and comfortable literally, while s/he is angry and aggressive actually. We take an sarcastic utterance as an example, e.g., “*Good idea, sit with her. Hold her, comfort her. And if the moment feels right,*

see if you can cop a feel.”. CFN first mistakenly treats it as a positive sentiment utterance, and thus feeds this wrong sentiment identification into the complex-valued embedding, then makes a wrong decision to regard it as a non-sarcastic utterance. Further, we also notice that a small amount of errors occur when an utterance expresses very negative sentiment. CFN may mix negative sentiment up with sarcasm polarity. This is due to the subtle difference between sentiment and sarcasm analysis. Discriminating irony attitude from negative sentiment is a tricky and complex problem, which is an open area of research. Table VI shows a few misclassification cases.

VI. Conclusions and Future Work

Sarcasm detection in conversation is an interesting and challenging AI task. In this paper, we introduce quantum theory and complex world into classical sarcasm detection. We propose a complex-valued fuzzy network to capture both the vagueness and uncertainty of human language in sarcastic expression. The main idea is to treat the utterance as a quantum superposition of a set of separate words, use the complex-valued vector to represent it. The contextual dependency among utterances is described as the interaction between a quantum system and its surrounding environment, which can be computed by the tensor product of individual utterance. Then, the speaker’s sarcastic attitude is viewed as a quantum mixed system composed of composite systems, which is mathematically encapsulated in a density matrix. A fuzzy quantum measurement is performed on the density matrix of each target utterance to yield the probabilistic outcomes. The experimental results show that our proposed CFN model largely outperforms a number of strong sarcasm detection methods. To the best of our knowledge, this work is the first that brings together quantum theory and fuzzy logic for SDC.

Fuzzy logic provides us with a means to deal with vagueness and uncertainty, and quantum logic may give us more insights into the semantics behind the fuzzy norms algebraic product and algebraic sum. Since there are closely relationships between quantum theory and fuzzy logic, Future works will bridge them together and design a quantum fuzzy neural network for sarcasm detection. We also plan to provide a description of quantum mechanics in terms of a deterministic fuzziness using quantum fuzzy model. Moreover, since human language usually involves multimodal records, e.g., image, video, audio, etc., future works will also focus on incorporating multimodal information into the CFN model and designing a multimodal sarcasm detection model based on fuzzy logic and quantum theory. At the heart of this multimodal model is how to use fuzzy logic to represent visual and acoustic features.

References

- [1] R. W. Gibbs, "On the psycholinguistics of sarcasm." *Journal of experimental psychology: general*, vol. 115, no. 1, p. 3, 1986.
- [2] S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, and S. Poria, "Towards multimodal sarcasm detection (an obviously perfect paper)," in *ACL 2019 : The 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 4619–4629.
- [3] L. Chen, W. Su, M. Wu, W. Pedrycz, and K. Hirota, "A fuzzy deep neural network with sparse autoencoder for emotional intention understanding in human-robot interaction," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 7, pp. 1252–1264, 2020.
- [4] Y. Cai, H. Cai, and X. Wan, "Multi-modal sarcasm detection in twitter with hierarchical fusion model," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 2506–2515.
- [5] D. S. Chauhan, S. Dhanush, A. Ekbal, and P. Bhattacharyya, "Sentiment and emotion help sarcasm? a multi-task learning framework for multimodal sarcasm, sentiment and emotion analysis," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 4351–4360.
- [6] R. A. Potamias, G. Siolas, and A.-G. Stafylopatis, "A transformer-based approach to irony and sarcasm detection," *Neural Computing and Applications*, pp. 1–12, 2020.
- [7] D. G. Maynard and M. A. Greenwood, "Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis," in *Lrec 2014 proceedings*. ELRA, 2014.
- [8] D. S. Chauhan, S. Dhanush, A. Ekbal, and P. Bhattacharyya, "All-in-one: A deep attentive multi-task learning framework for humour, sarcasm, offensive, motivation, and sentiment on memes," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2020, pp. 281–290.
- [9] V. Haripriya and P. G. Patil, "Detection of sarcasm from consumer sentiments on social media about luxury brands," in *International Conference on Recent Trends in Image Processing and Pattern Recognition*. Springer, 2018, pp. 652–667.
- [10] S. Rendalkar and C. Chandankhede, "Sarcasm detection of online comments using emotion detection," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2018, pp. 1244–1249.
- [11] Y. Zhang, P. Tiwari, D. Song, X. Mao, P. Wang, X. Li, and H. M. Pandey, "Learning interaction dynamics with an interactive lstm for conversational sentiment analysis," *Neural Networks*, vol. 133, pp. 40–56, 2021.
- [12] H. Gregory, S. Li, P. Mohammadi, N. Tarn, R. Draelos, and C. Rudin, "A transformer approach to contextual sarcasm detection in twitter," in *Proceedings of the Second Workshop on Figurative Language Processing*, 2020, pp. 270–275.
- [13] G. W. F. Hegel, *The philosophy of fine art*. G. Bell and sons, Limited, 1920, vol. 3.
- [14] A. Goli, E. B. Tirkolaee, and N. S. Aydin, "Fuzzy integrated cell formation and production scheduling considering automated guided vehicles and human factors," *IEEE Transactions on Fuzzy Systems*, 2021.
- [15] A. Goli and B. Malmir, "A covering tour approach for disaster relief locating and routing with fuzzy demand," *International Journal of Intelligent Transportation Systems Research*, vol. 18, no. 1, pp. 140–152, 2020.
- [16] E. B. Tirkolaee, A. Goli, A. Faridnia, M. Soltani, and G.-W. Weber, "Multi-objective optimization for the reliable pollution-routing problem with cross-dock selection using pareto-based algorithms," *Journal of Cleaner Production*, vol. 276, p. 122927, 2020.
- [17] M. Yang, S. Liu, K. Chen, H. Zhang, E. Zhao, and T. Zhao, "A hierarchical clustering approach to fuzzy semantic representation of rare words in neural machine translation," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 992–1002, 2020.
- [18] C. Liu, L. Song, J. Zhang, K. Chen, and J. Xu, "Self-supervised learning for specified latent representation," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 1, pp. 47–59, 2019.
- [19] Y. Zhang, D. Song, X. Li, P. Zhang, P. Wang, L. Rong, G. Yu, and B. Wang, "A quantum-like multimodal network framework for modeling interaction dynamics in multiparty conversational sentiment analysis," *Information Fusion*, vol. 62, pp. 14–31, 2020.
- [20] A. Sordani, J.-Y. Nie, and Y. Bengio, "Modeling term dependencies with quantum language models for ir," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013, pp. 653–662.
- [21] P. Zhang, J. Niu, Z. Su, B. Wang, L. Ma, and D. Song, "End-to-end quantum-like language models with application to question answering," pp. 5666–5673, 2018.
- [22] Y. Zhang, D. Song, X. Li, and P. Zhang, "Unsupervised sentiment analysis of twitter posts using density matrix representation," in *European Conference on Information Retrieval*. Springer, 2018, pp. 316–329.
- [23] Y. Zhang, D. Song, P. Zhang, X. Li, and P. Wang, "A quantum-inspired sentiment representation model for twitter sentiment analysis," *Applied Intelligence*, vol. 49, no. 8, pp. 3093–3108, 2019.
- [24] Q. Li, B. Wang, and M. Melucci, "Cnm: An interpretable complex-valued network for matching," in *NAACL-HLT (1)*, 2019, pp. 4139–4148.
- [25] I. Schmitt, A. Nürnberger, and S. Lehrack, "On the relation between fuzzy and quantum logic," in *Views on Fuzzy Sets and Systems from Different Perspectives*. Springer, 2009, pp. 417–438.
- [26] S. K. Bharti, K. S. Babu, and S. K. Jena, "Parsing-based sarcasm sentiment recognition in twitter data," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 2015, pp. 1373–1380.
- [27] I. Hernández-Farías, J.-M. Benedí, and P. Rosso, "Applying basic features from sentiment analysis for automatic irony detection," in *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 2015, pp. 337–344.
- [28] M. Bouazizi and T. O. Ohtsuki, "A pattern-based approach for sarcasm detection on twitter," *IEEE Access*, pp. 5477–5488, 2016.
- [29] P. Clews and J. Kuzma, "Rudimentary lexicon based method for sarcasm detection," *International Journal of Academic Research and Reflection*, vol. 5, no. 4, pp. 24–33, 2017.
- [30] S. Hiai and K. Shimada, "A sarcasm extraction method based on patterns of evaluation expressions," in *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, 2016, pp. 31–36.
- [31] M. Abulaish and A. Kamal, "Self-deprecating sarcasm detection: an amalgamation of rule-based and machine learning approach," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 2018, pp. 574–579.
- [32] E. Lunando and A. Purwarianti, "Indonesian social media sentiment analysis with sarcasm detection," in *2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, 2013, pp. 195–198.
- [33] T. Ptáček, I. Habernal, and J. Hong, "Sarcasm detection on czech and english twitter," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 213–223.
- [34] S. Mukherjee and P. K. Bala, "Sarcasm detection in microblogs using naïve bayes and fuzzy clustering," *Technology in Society*, vol. 48, pp. 19–27, 2017.
- [35] A. Kumar and H. Sharma, "Prod: A potential rumour origin detection model using supervised machine learning," in *International Conference on Intelligent Computing and Smart Communication 2019*. Springer, 2020, pp. 1269–1276.
- [36] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm detection using multi-head attention based bidirectional lstm," *IEEE Access*, vol. 8, pp. 6388–6397, 2020.
- [37] A. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm detection on twitter: A behavioral modeling approach," in *Proceedings of the eighth ACM international conference on web search and data mining*, 2015, pp. 97–106.
- [38] A. Joshi, V. Tripathi, P. Bhattacharyya, and M. Carman, "Harnessing sequence labeling for sarcasm detection in dialogue from tv series 'friends'," in *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, 2016, pp. 146–155.
- [39] S. Belen and C. Pearce, "Rumours with general initial conditions," *The ANZIAM Journal*, vol. 45, no. 3, pp. 393–400, 2004.
- [40] S. Belen, C. Y. Kaya, and C. Pearce, "Impulsive control of rumours with two broadcasts," *The ANZIAM Journal*, vol. 46, no. 3, pp. 379–391, 2005.
- [41] S. Belen, T. Langtry, and S. A. UTS, "Computational and numerical aspects of repeated rumours," pp. 1339–1348, 2003.
- [42] S. Belen, E. Kropat, and G.-W. Weber, "On the classical maki-thompson rumour model in continuous time," *Central European Journal of Operations Research*, vol. 19, no. 1, pp. 1–17, 2011.
- [43] G.-W. Weber, B. Gürbüz, H. Mawengkang, and I. Husein, "Numerical approach for rumor propagation model," pp. 339–348, 08 2019.
- [44] —, "Rumor propagation model: Or and numerical approach," pp. 120–126, 2019.
- [45] B. Gürbüz and G.-W. Weber, "A numerical scheme for solving a rumor propagation model," pp. 30–36, 06 2019.

- [46] S. Poria, E. Cambria, D. Hazarika, and P. Vij, "A deeper look into sarcastic tweets using deep convolutional neural networks," *arXiv preprint arXiv:1610.08815*, 2016.
- [47] M. Zhang, Y. Zhang, and G. Fu, "Tweet sarcasm detection using deep neural network," in *Proceedings of COLING 2016, The 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 2449–2460.
- [48] I. Chaturvedi, R. Satapathy, S. Cavallari, and E. Cambria, "Fuzzy commonsense reasoning for multimodal sentiment analysis," *Pattern Recognition Letters*, vol. 125, pp. 264–270, 2019.
- [49] S. Vashishtha and S. Susan, "Fuzzy rule based unsupervised sentiment analysis from social media posts," *Expert Systems with Applications*, vol. 138, p. 112834, 2019.
- [50] N. Chatterjee, T. Aggarwal, and R. Maheshwari, "Sarcasm detection using deep learning-based techniques," in *Deep Learning-Based Approaches for Sentiment Analysis*. Springer, 2020, pp. 237–258.
- [51] L. Liu, J. L. Priestley, Y. Zhou, H. E. Ray, and M. Han, "A2text-net: A novel deep neural network for sarcasm detection," in *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*. IEEE, 2019, pp. 118–126.
- [52] N. Jaiswal, "Neural sarcasm detection using conversation context," in *Proceedings of the Second Workshop on Figurative Language Processing*, 2020, pp. 77–82.
- [53] D. Ghosh, A. R. Fabbri, and S. Muresan, "The role of conversation context for sarcasm detection in online interactions," in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, 2017, pp. 186–196.
- [54] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, "Sentiment and sarcasm classification with multitask learning," *IEEE Intelligent Systems*, vol. 34, no. 3, pp. 38–43, 2019.
- [55] A. Goli, H. K. Zare, R. Tavakkoli-Moghaddam, and A. Sadegheih, "Multiobjective fuzzy mathematical model for a financially constrained closed-loop supply chain with labor employment," *Computational Intelligence*, vol. 36, no. 1, pp. 4–34, 2020.
- [56] A. Goli, H. K. Zare, R. Tavakkoli-Moghaddam, and A. Sadegheih, "Hybrid artificial intelligence and robust optimization for a multi-objective product portfolio problem case study: The dairy products industry," *Computers & industrial engineering*, vol. 137, p. 106090, 2019.
- [57] E. Kropat, G.-W. Weber, and B. Akteke, "Eco-finance networks under uncertainty," *International Conference on Engineering Optimization*, pp. 1–13, 2008.
- [58] E. Kropat, G.-W. Weber, and C. S. Pedamallu, "Regulatory networks under ellipsoidal uncertainty—data analysis and prediction by optimization theory and dynamical systems," in *Data Mining: Foundations and Intelligent Paradigms*. Springer, 2012, pp. 27–56.
- [59] E. Kropat, G.-W. Weber, and J.-J. Rückmann, "Regression analysis for clusters in gene-environment networks based on ellipsoidal calculus and optimization," *Dynamics of Continuous, Discrete and Impulsive Systems, Series B: Applications & Algorithms*, vol. 17, no. 5, pp. 639–657, 2010.
- [60] E. Kropat, A. Özmen, G.-W. Weber, S. Meyer-Nieberg, and O. Deftlerli, "Fuzzy prediction strategies for gene-environment networks—fuzzy regression analysis for two-modal regulatory systems," *RAIRO-Operations Research*, vol. 50, no. 2, pp. 413–435, 2016.
- [61] A. Özmen, E. Kropat, and G.-W. Weber, "Spline regression models for complex multi-modal regulatory networks," *Optimization Methods and Software*, vol. 29, no. 3, pp. 515–534, 2014.
- [62] E. Kropat and G. W. Weber, "Fuzzy target-environment networks and fuzzy-regression approaches," *Numerical Algebra, Control & Optimization*, vol. 8, no. 2, p. 135, 2018.
- [63] J. Li, Z. Wang, Y. Shen, and Y. Wang, "Interval observer design for discrete-time uncertain takagi–sugeno fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 4, pp. 816–823, 2019.
- [64] G. Cattaneo, M. L. D. Chiara, and R. Giuntini, "Fuzzy intuitionistic quantum logics," *Studia Logica*, vol. 52, no. 3, pp. 419–442, 1993.
- [65] A. Granik and H. J. Caulfield, "Fuzziness in quantum mechanics," *Physics Essays*, vol. 9, no. 3, pp. 496–505, 1996.
- [66] B. Wang, Q. Li, M. Melucci, and D. Song, "Semantic hilbert space for text representation learning," in *The World Wide Web Conference*, 2019, pp. 3293–3299.
- [67] L. Masanes and M. P. Müller, "A derivation of quantum theory from physical requirements," *New Journal of Physics*, vol. 13, no. 6, p. 63001, 2011.
- [68] D. Ghosh, A. Vajpayee, and S. Muresan, "A report on the 2020 sarcasm detection shared task," in *Proceedings of the Second Workshop on Figurative Language Processing*, 2020, pp. 1–11.
- [69] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT 2019: Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2019, pp. 4171–4186.
- [70] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.
- [71] Y. Zhang, Q. Li, D. Song, P. Zhang, and P. Wang, "Quantum-inspired interactive networks for conversational sentiment analysis," in *IJCAI2019*, 2019, pp. 5436–5442. [Online]. Available: <https://academic.microsoft.com/paper/2963533390>
- [72] S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L.-P. Morency, "Context-dependent sentiment analysis in user-generated videos," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 873–883. [Online]. Available: <https://www.aclweb.org/anthology/P17-1081>
- [73] A. Kumar Jena, A. Sinha, and R. Agarwal, "C-net: Contextual network for sarcasm detection," in *Proceedings of the Second Workshop on Figurative Language Processing*. Online: Association for Computational Linguistics, Jul. 2020, pp. 61–66. [Online]. Available: <https://www.aclweb.org/anthology/2020.figlang-1.8>